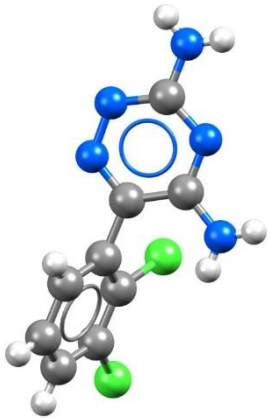
Introduction to ConQuest (CQ-001)

Developed using
2023.1 CSD Release
CSD version 5.44 (April 2023)

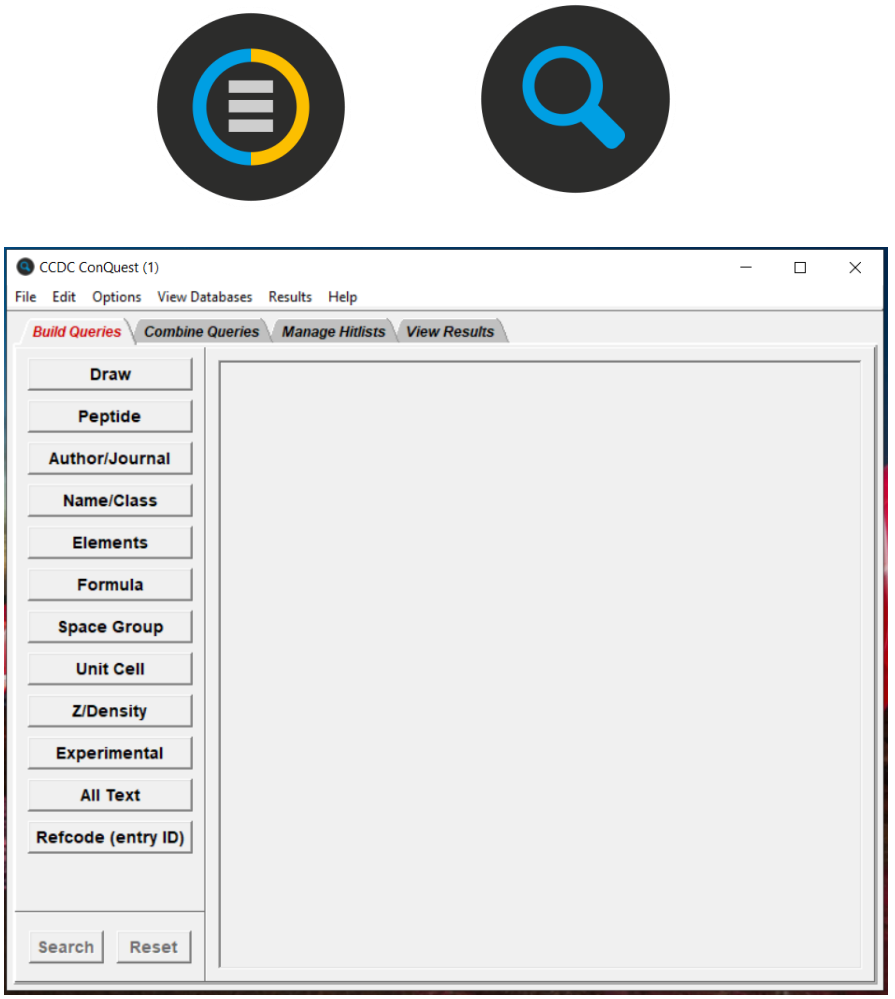


CCDC
advancing structural science

Table of Contents

Introduction

ConQuest is the desktop search interface to the Cambridge Structural Database (CSD). All textual, numeric and structural data stored within the CSD can be searched using ConQuest. ConQuest provides an extensive range of flexible search options including searching based on compound name, formula, elemental composition, and literature references to name a few. Chemical *substructure* search in ConQuest also includes the ability to define chemical constrains such as charge, hybridization state and cyclicity.



Before beginning this tutorial, ensure that you have a registered copy of CSD-Core or CSD-Enterprise installed on your computer. Please contact your site administrator or workshop host for further information.

Learning Outcomes

This workshop will guide you through using ConQuest for *substructure* searching. At the end of this tutorial, you will be able to:

- Draw a 2D structure in ConQuest and conduct a basic search.
- Search for similar molecules.
- Combine queries and search results.
- Analyse and visualize results.
- Export search results.

This workshop will take approximately 1.5 hours to complete. The words in *Blue Italics* are reported in the Glossary at the end of this handout.

Pre-required Skills

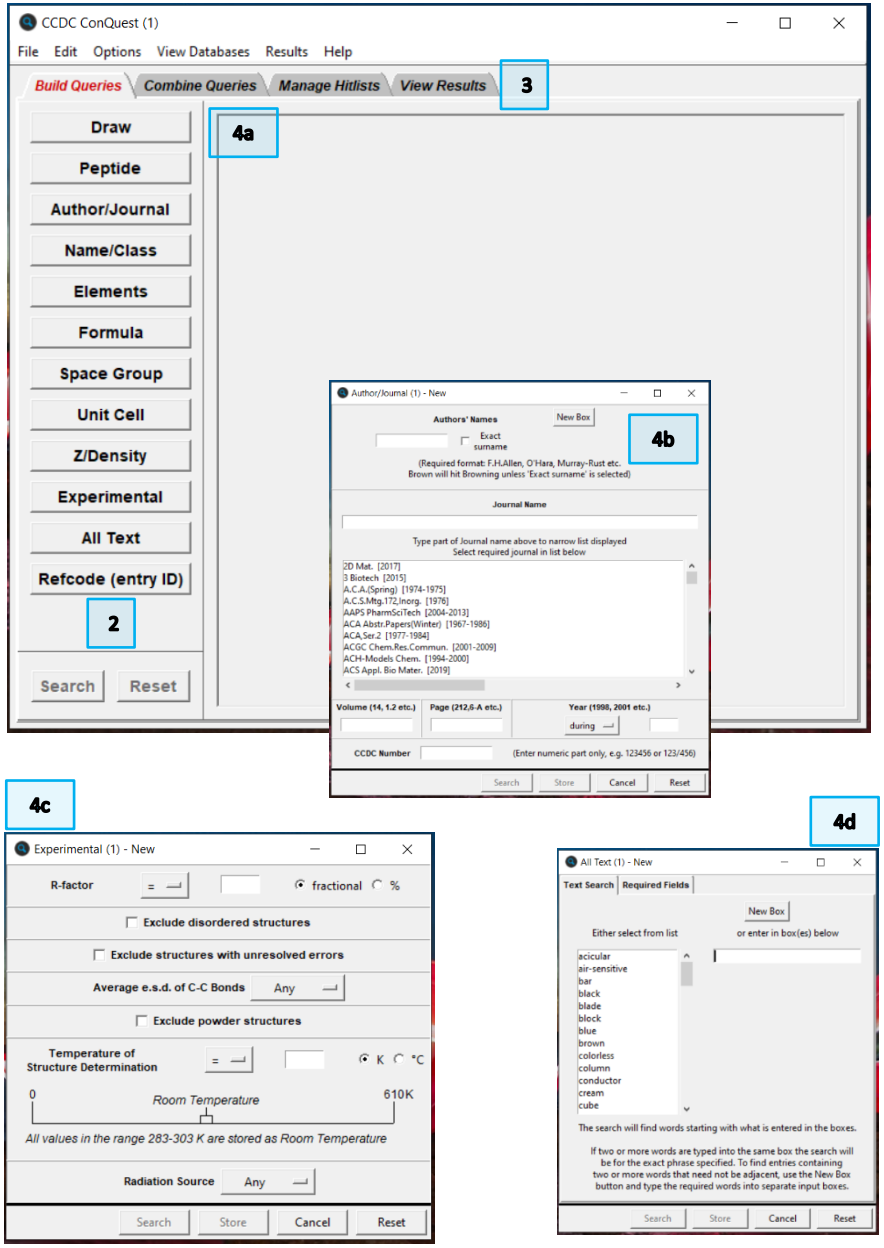
There are no pre-required skills for this workshop.

Materials

There are no additional materials required for this workshop.

Overview of ConQuest

1. Launch ConQuest by clicking the ConQuest Icon on your desktop or launching it from the Start or Applications menu.
2. The ConQuest main window shows all the search routines you can perform on the left-hand side of the window.
3. The row of tabs across the top of the window will guide you through the steps of the search process.
4. Some example searches are:
    a. Draw – substructure and 3D information searching.
    b. Author/Journal – bibliographic searching.
    c. Experimental – experimental set up searching.
    d. All Text – generic text-based searching.
5. The majority of the searching we will do in these tutorials will be substructure searching, so we will focus on the Draw tab here.

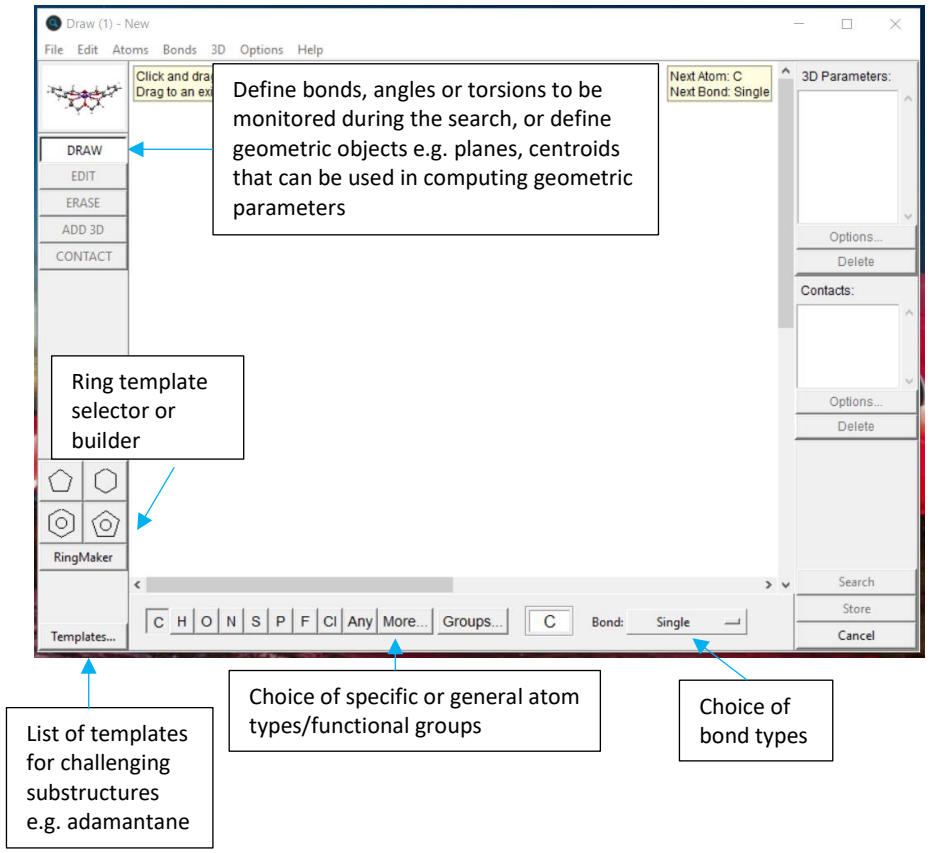


Introduction to the Draw Window

All drawing takes place in the central white area of the *Draw* window. In addition to creating 2D chemical structure sketches, the *Draw* window allows for the inclusion of 3D parameters for searching or for filtering.

ConQuest sketching conventions.

- Left click in the sketcher to insert the selected atom type.
- Left click and drag to sketch two bonded atoms.
- Use the **Edit** button to modify properties of or delete atoms, bonds or entire substructures.
- Right-click on atoms or bonds to modify their properties.
- Use the **Templates…** button to pick from a list of CSD editor devised and drawn substructures.
- Use the **More…** button to find less frequently used element types, or generic atom type groups (e.g. halogens), or define custom element combinations (e.g. C or N or O).



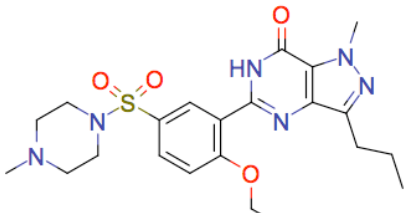
Example 1. Searching for Similar Molecules

Sildenafil citrate or “Viagra” is a well-known drug. Since sildenafil’s release in 1998, other competitor products have been released, e.g. vardenafil, that are structurally similar but have subtle chemical differences (see right).

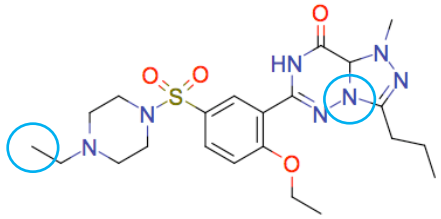
Are there other entries in the CSD that contain similar fused ring substructures to those observed in both sildenafil and vardenafil?

This tutorial will take you through the steps needed to search the CSD for such similar compounds**. You will learn how to sketch and edit a fragment and how to view your results.**

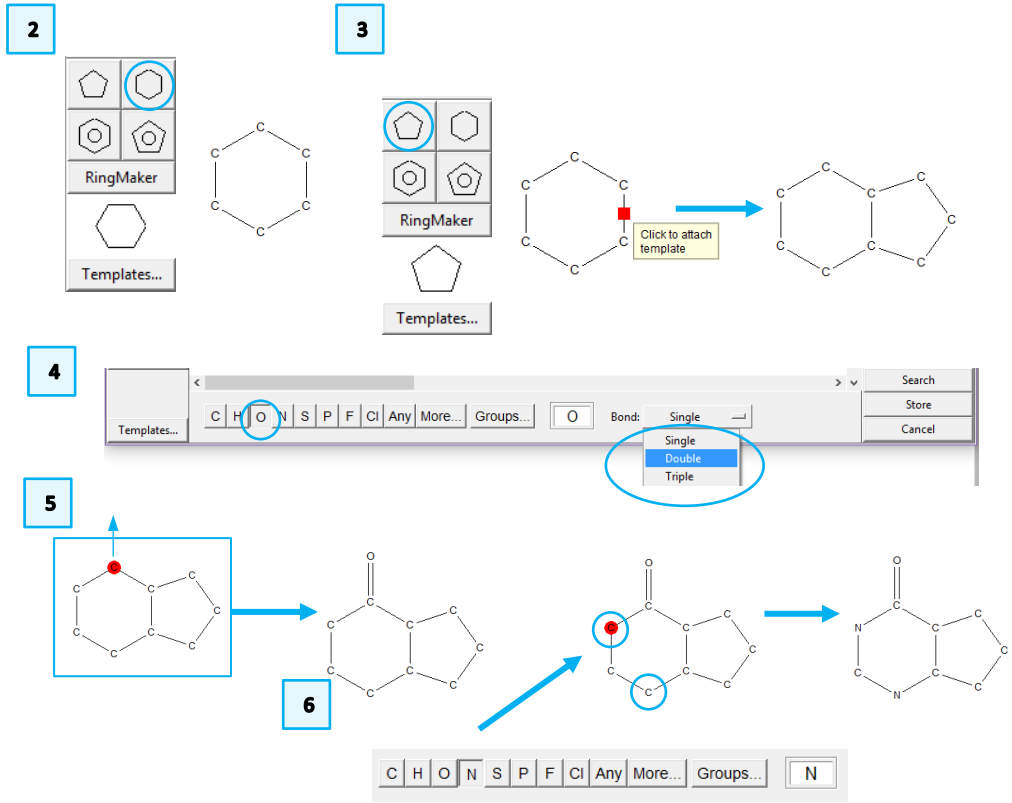
Sildenafil (neutral form of active ingredient)



Vardenafil (neutral form of active ingredient)



1. Launch ConQuest and open the sketcher by clicking the **Draw** button.
2. Start to sketch the sildenafil substructure shown on the right by first clicking on a six-membered ring template and clicking in the sketcher area.
3. Now click on a five-membered ring template, hover over the right-hand C-C bond to attach the five-membered ring to the six-membered ring.
4. Add the carbonyl oxygen to the substructure by first selecting the O atom as shown below, and then changing *Single* to *Double* in the **Bond** pull-down menu.
5. Click on the C atom to which the O atom should be bonded and then drag upwards while holding the left mouse button down.
6. Introduce two nitrogen atoms into the six-membered ring at the positions shown by left-clicking the N button and then left-clicking on the atoms to be modified.



7. Add variable atom types (C or N) to the five-membered ring. Select *Other Elements* from the **More** drop-down menu, then activate the Multi Pick radio button and select **C** and **N** from the periodic table. Click OK to store selection.

8. A new QA atom type has been produced. Select the atoms shown and modify these to be QA (i.e. either C or N) by left clicking them.

9. Add a double bond to the six-membered ring shown by ensuring *Double* is selected from the **Bond** drop-down menu, then left-clicking on the bond.

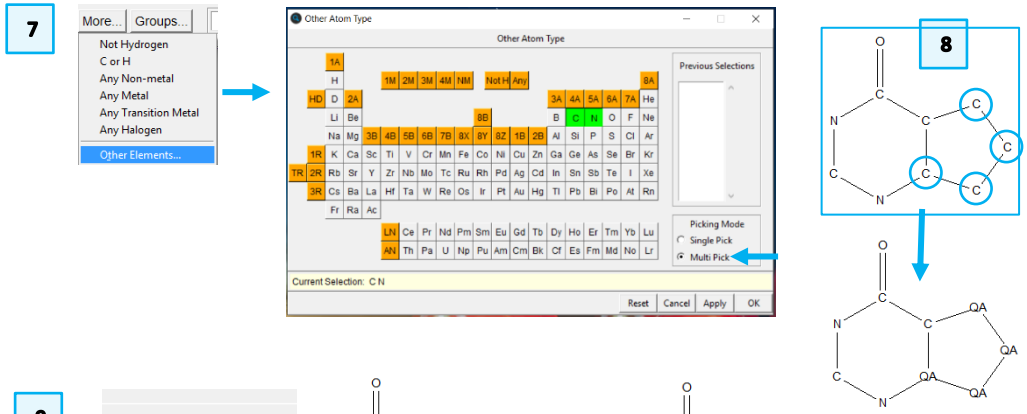
10. Modify the bond types shown. Click the **Bond** type drop-down menu and select *Variable*. Tick the boxes for *Single* and *Double* as shown, then click **OK.**

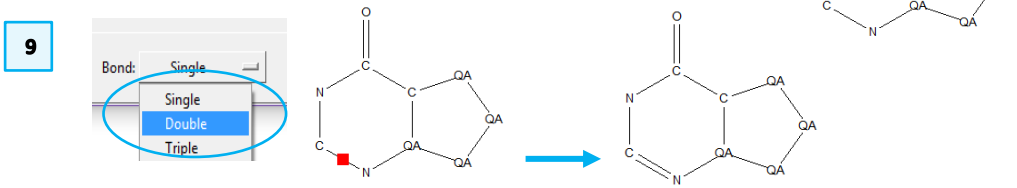
11. Now click on all the bonds in the five-membered ring to change them to the variable 1, 2 bond type.

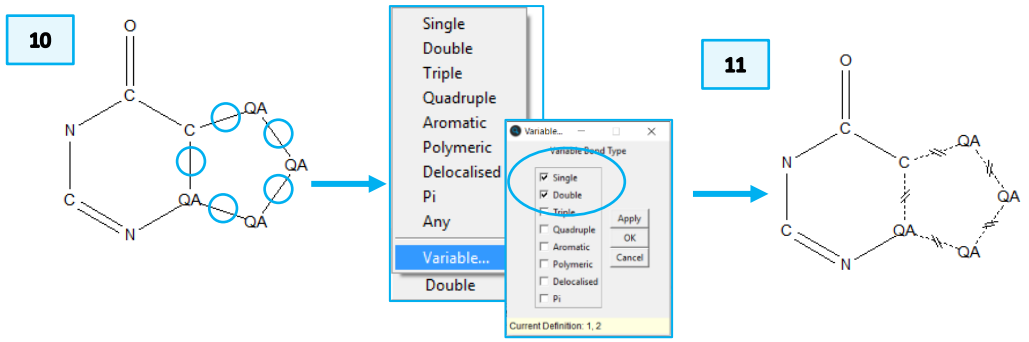
12. Exclude any further ring fusion by right-clicking on the N atom adjacent to the carbonyl, clicking on *Hydrogens* from the drop down and selecting *1* from the list.

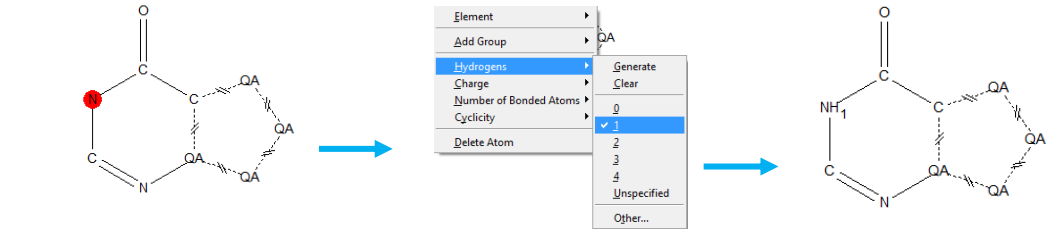
13. Now that you have finished sketching the fragment, start the search by clicking the **Search** button.

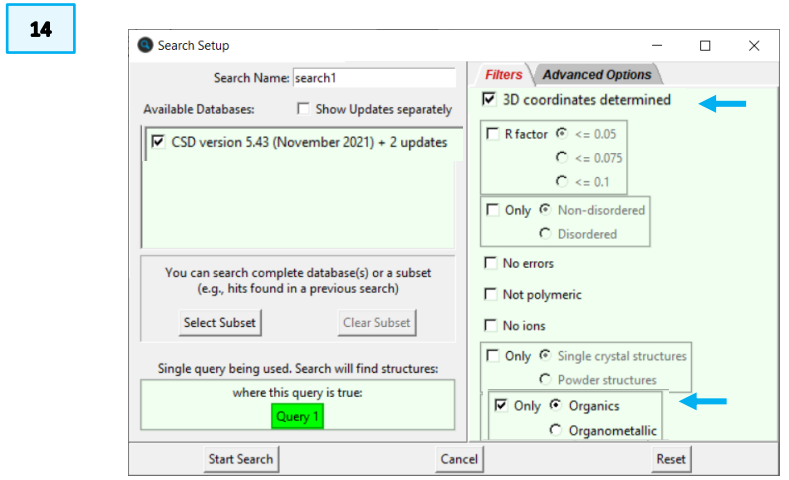
14. Tick the boxes for the "3D coordinates determined" and "Only Organics" filters in the *Search Setup* window, then click **Start Search.**











Viewing search results

15. All data stored with each CSD entry can be accessed from the tabs on the left, e.g. publication and experimental details.

16. The matched substructure is highlighted in red in the 2D diagram.

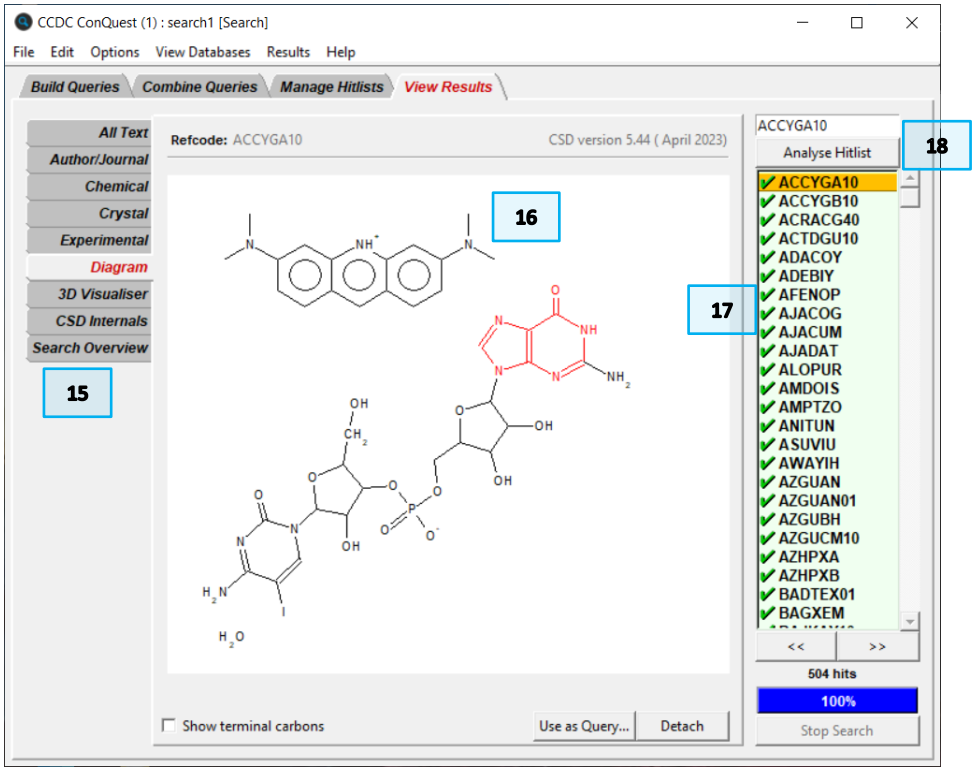
17. *Hitlist* of CSD entries that contain the matched substructure are shown on the right side of the window. There should be at least 504[1] structures returned (The numbers may vary depending on the version of the database).

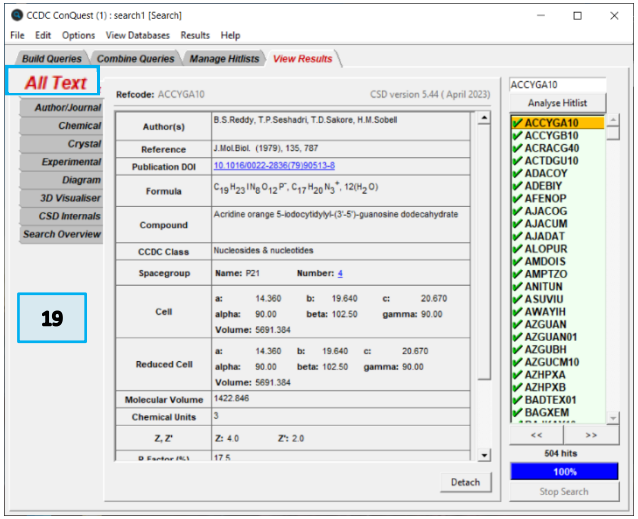
18. To inspect crystal packing, select **Analyse Hitlist** > **Visualise Structures**. This will open your results in Mercury; however, we are not exploring this yet.

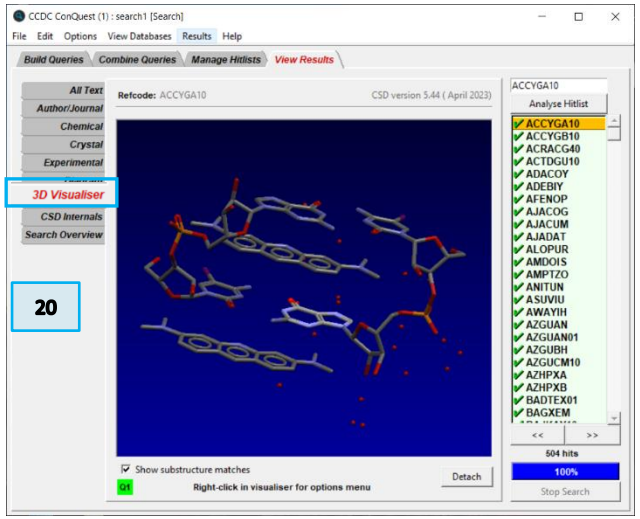
19. The *All Text* tab displays all textual information stored with the CSD entry.

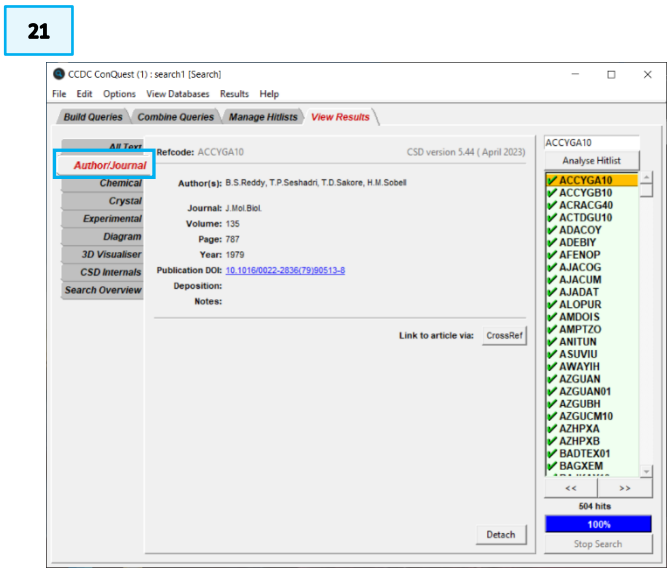
20. Structures can also be viewed in 3D using the *3D Visualiser* tab.

21. The *Author/Journal* tab shows publication details including links to the manuscript DOI where available.









[1] As of version 2023.1 of the CSD.

Conclusions

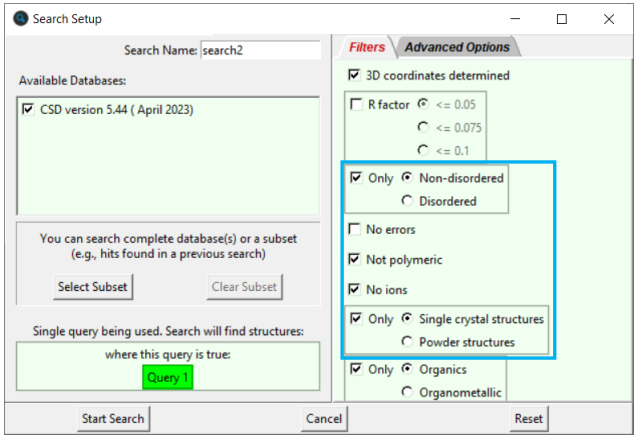
There are at least 504 CSD entries that contain similar substructures to the five- and six-membered fused ring core found in sildenafil and vardenafil. Sildenafil itself appears in the search results (CSD *refcode* QEGTUT), vardenafil, however, is not in the CSD. Two closely related compounds, BITZOT and PEYGIL, are. Many other hits from this search have drug-like activity or are natural products.

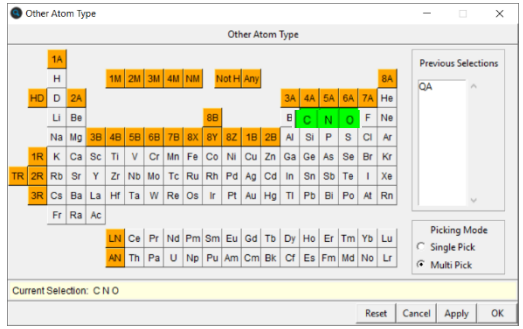
In the case of the five-membered ring, the bond and atom types in the substructure were left variable (single and double, and C, N, respectively). These could be further refined, for instance to match exactly the atom types and bonding found in sildenafil if additional specificity were required.

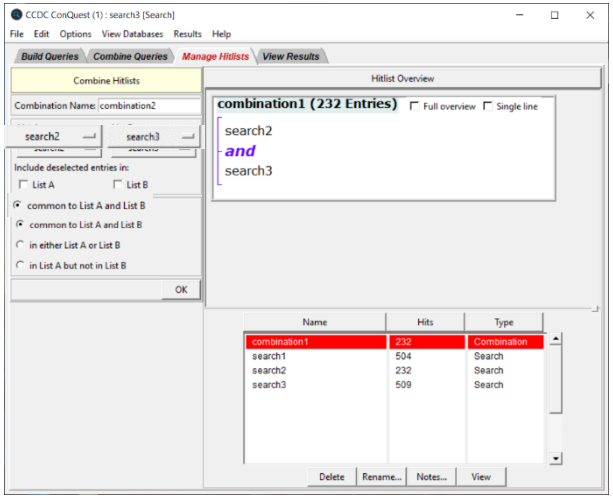
You should now be familiar with sketching substructures using the ConQuest *Draw* window; how to change bond and atom types; how to use ring templates; and how to view the data for each hit in your results list.

Exercises

- Try changing the filters to exclude *disordered*, ionic, polymeric and powder structures as well as those with errors. How do your results change? (***Should return at least 232 hits***)
- Change the 'QA' atom type to include O as well as C and N and using the filters 3D coordinates determined and only organics, conduct a new search. How does this change your results? (***Should return at least 509 hits***)
- Can you design a similar search to focus on a different fragment of the sildenafil molecule? How many hits do you get?
- Explore the **Templates…** button in the **Draw** window to see what other useful templates are available.
- Try designing a substructure search for your own compounds of interest.
- Hint: Can you use the *Manage Hitlists* tab to combine hitlists to find common molecules between the above search results?







Example 2. More Searching in ConQuest

We can demonstrate the effect small changes on our search fragment have on the results of our searches by looking at the series of alcohols shown on the right.

This example will demonstrate additional sketching techniques, combining different queries with Boolean operators, and exporting your results.

1. Launch ConQuest and open the sketcher by clicking the **Draw** button.

2. Sketch a simple alcohol fragment by first adding a carbon atom and then adding an oxygen atom to it with a single bond.

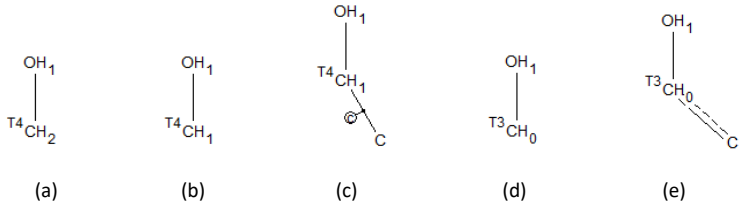
3. Right click the O, choose *Hydrogens* and then *1* from the drop-down menu to add the H atom.

4. To ensure that the carbon atom is bonded to at least 2 hydrogen atoms and one other atom in addition to the OH group, first add two H atoms as in step 3 above. Then right-click the C again and choose *Number of Bonded Atoms*, and then *4* from the drop-down menu. This will produce fragment "a" from the list at the right.

5. Click the **Search** button, tick the boxes for *3D coordinates determined* and *Only Organics* in the **Search Setup** window. Then click the **Start Search** button.

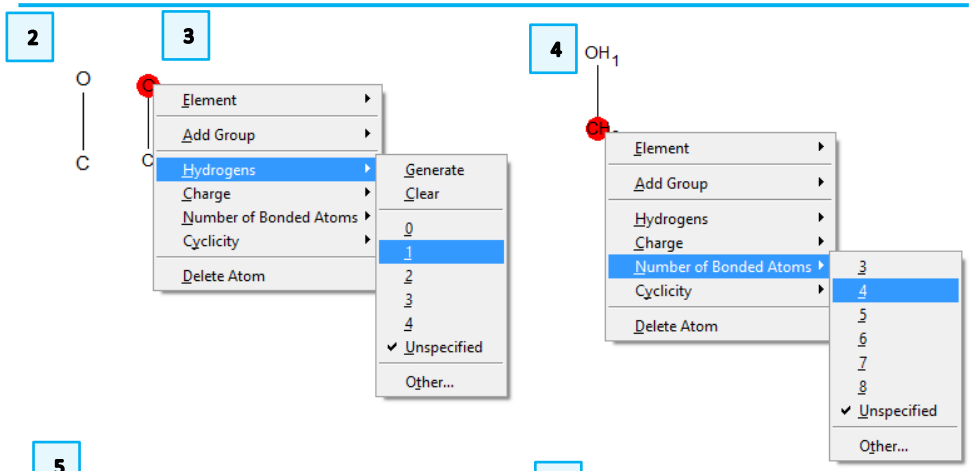
6. You should get over 17,700 hits with this query. Use the arrow keys to scroll through the *refcode* list to view the results. *What do you find*?

7. When you are done looking at your results, click the *Build Queries* tab to return to the Query window. You should see your first search fragment listed as Query 1 as in Step 5 to the right.



These subtly different definitions would retrieve the following types of functional group:

(a) primary alcohols
(b) cyclic and acylic secondary alcohols
(c) cyclic secondary alcohols only
(d) unsaturated cyclic and acyclic alcohols and carboxylic acids
(e) unsaturated cyclic alcohols only e.g. phenols



8. Click the **Edit…** button in the query window to return to the **Draw** window.

9. Right-click on the C atom and choose *Hydrogens,* and *1* from the drop-down menu as in step 3 above. Leave the *Number of Bonded Atoms* set to 4 to match a secondary alcohol (fragment "b" above).

10. Click **Search**. A pop-up warning will ask if you want to overwrite the existing Query. Click **No** to save this as a new Query.

11. Make sure the "3D coordinates determined" and "Only Organics" filters are ticked and start the search.

12. You should get over 32,000 hits with this new query. Use the arrow keys to scroll through the *refcode list* to view the results. Notice how the hits are different from the first search.

13. Repeat steps 7 and 8 above to return to the **Draw** window with the fragment from Query 2 showing.

14. To draw a cyclic secondary alcohol, first draw another C bonded to the first one.

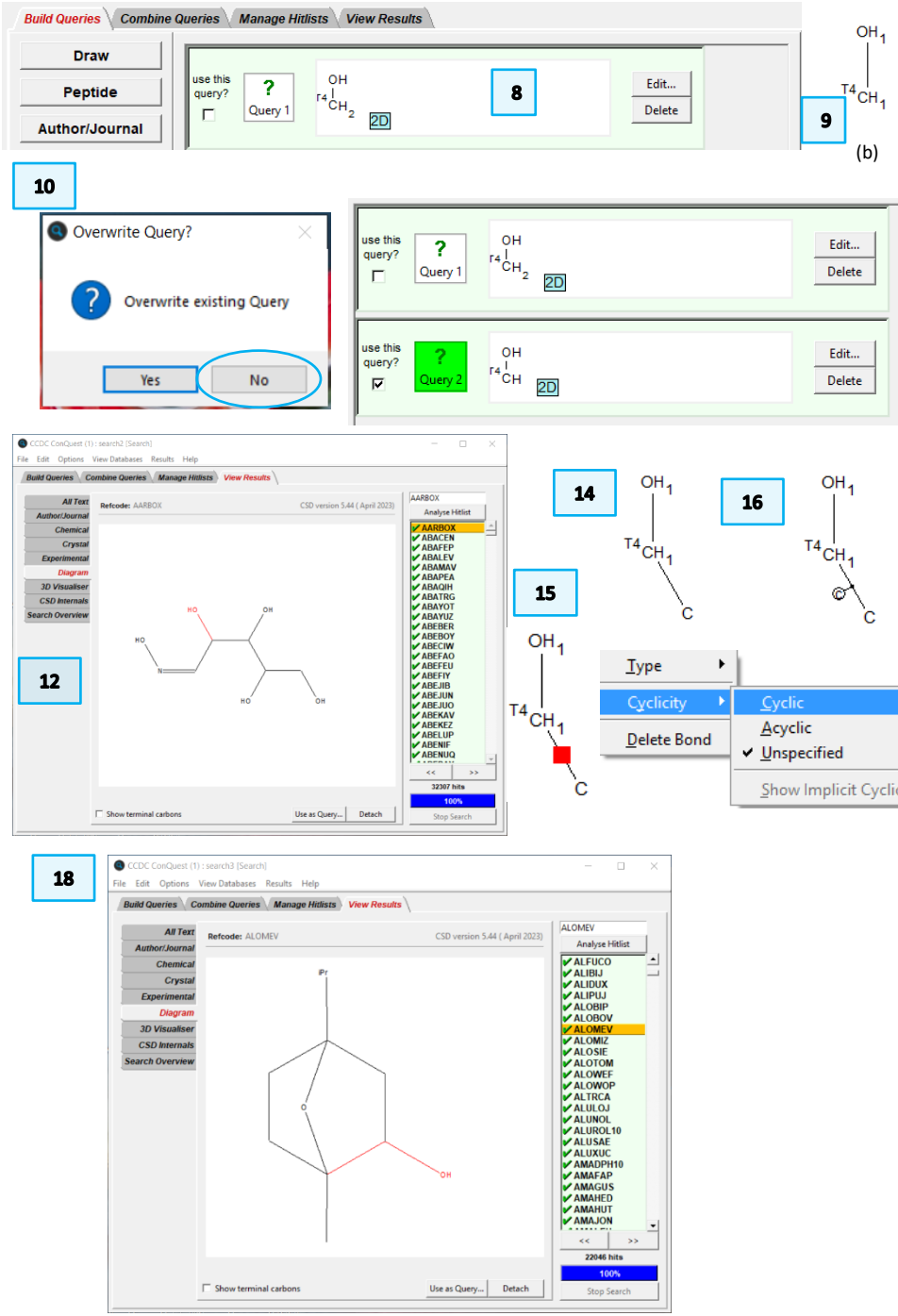
15. Now, right-click on the C-C bond, choose *Cyclicity* and *Cyclic* from the drop-down menu.

16. This will mark the bond as *cyclic* or belonging to a ring system (fragment C above).

17. Follow steps 10 and 11 above to save this as Query 3 and start the search.

18. This time you will get over 22,100 hits. Again, scroll through the *refcode list* to explore your results. Notice that refcode ALOMEV appears in both searches.

19. Again repeat steps 7 and 8 above to return to the **Draw** window showing the fragment from Query 2.



20. Right-click the C atom and choose *Hydrogens* and *0* from the drop-down menu. Right-click the C atom again and choose *Number of Bonded Atoms*, and then *3*. This will give you a fragment that matches unsaturated alcohols (fragment "d" above).

21. Click **Search** and then **No** to write this out as a new query.

22. Start the search again with "3D coordinates determined" and "Only Organics" filters set.

23. You should have over 65,000 hits for this query. Again, scroll through the refcode list to explore your results.

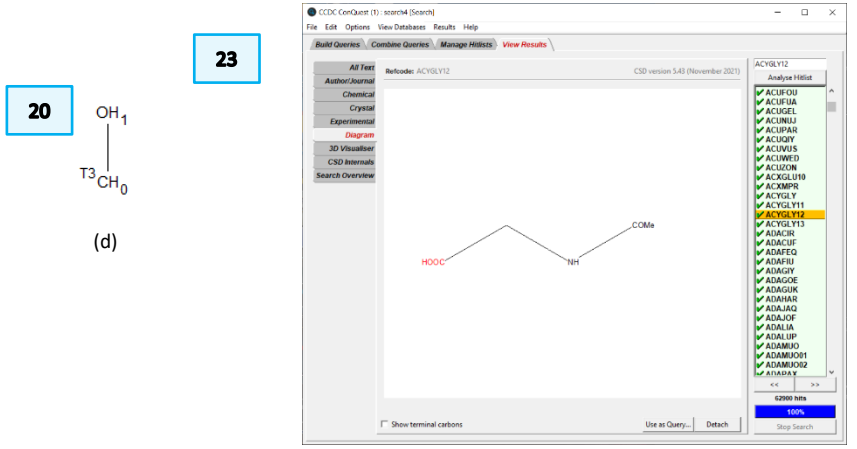
24. Finally, we want to search for a phenol alcohol (fragment "e"). Again, follow steps 7 and 8 above to edit Query 3.

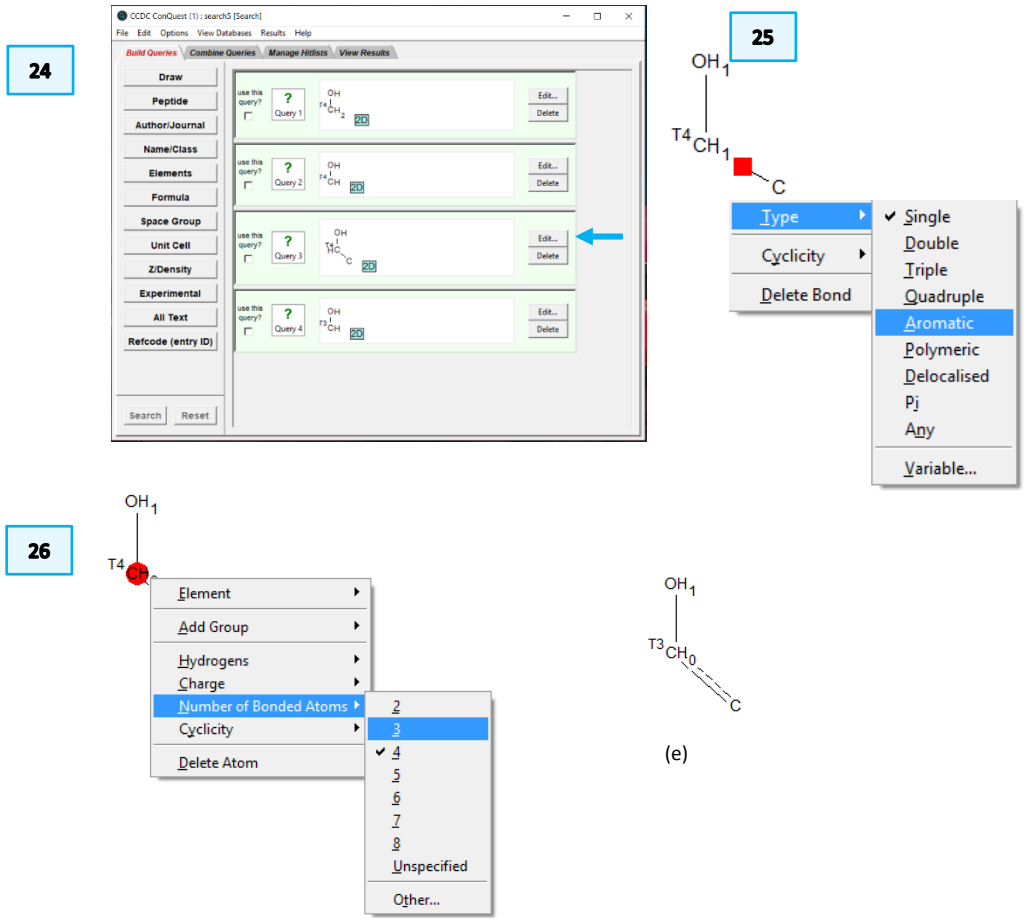
25. In the **Draw** window, right-click the single bond between the two carbons. Select *Cyclicity* and then *Unspecified* from the drop-down menu. Right-click the single bond again and choose *Type* and then *Aromatic* from the drop-down menu.

26. Right-click the central C atom, choose *Hydrogens* and then *0* from the drop-down menu. Then right-click the central C again, choose *Number of Bonded Atoms* and the *3* from the drop-down. This will set up the proper bonding for a phenol group.

27. Click **Search** and then **No** to save this fragment as a new query. Start the search with "3D coordinates determined" and "Only Organics" filters set.

28. You should now have over 34,000 hits for this query. Scroll through the refcode list to explore your results. Note that ALOLEU is included in this search as well as the previous search.





Combining Queries

1. Now you should have five separate queries representing the five different alcohol fragments above. You can check this by clicking on the *Combine Queries* tab of the window.

2. Let us look at finding compounds with both a cyclic secondary alcohol (Query 3) AND a phenol (Query 5). To do this, click the box that says "*Query 3*" and drag it to the top box labelled "*must have (boolean AND)*".

3. Do the same with the box that says "Query 5" so that you have both Query 3 and Query 5 in the top "must have" box.

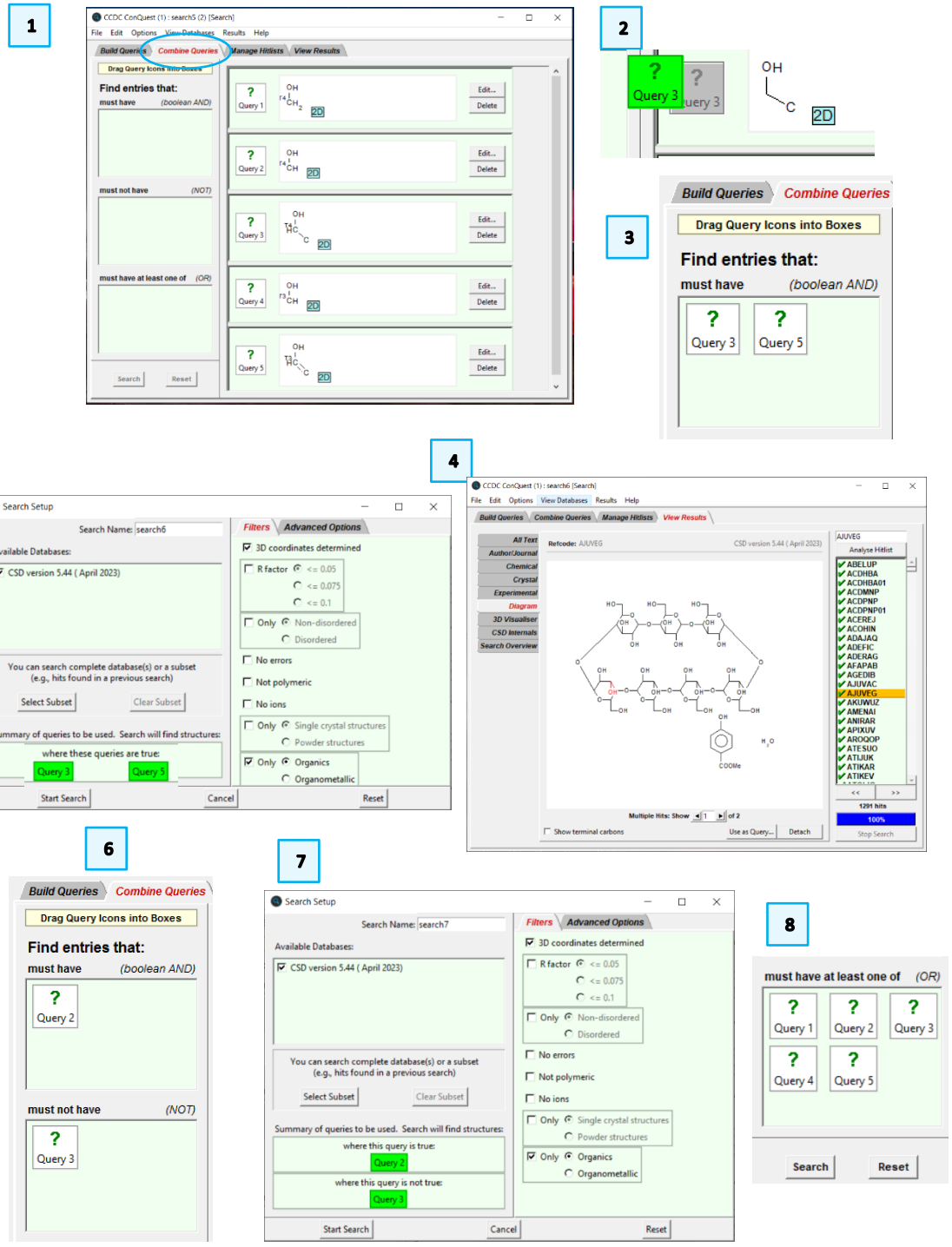
4. Click **Search** and make sure you have the "3D coordinates" and "Only Organics" filters ticked. You should get at least 1290 hits. Scroll through the refcode list to explore your results.

5. Query 2 returns all secondary alcohols while Query 3 returns only cyclic secondary alcohols. You can use the Boolean NOT operator to return only acyclic secondary alcohols.

6. Return to the *Combine Queries* tab. First, click and drag the Query 3 and Query 5 boxes back to the right side of the window (or click the **Reset** button). Then drag the Query 2 box to the "must have" box – we must have a secondary alcohol. Next drag the Query 3 box to the "must not have" box – we don't want any cyclic secondary alcohols.

7. Start the search as in step 4 above. You should obtain at least 10,200 hits. Scroll through your results to see that none of them contain cyclic secondary alcohols.

8. Finally, to find all compounds containing an alcohol group in this set, drag each Query box to the bottom "must have at least one of (OR)" box. Start the search as usual. It will return over 103,00 hits! Care must be taken when using this option as it returns any structure that meets at least one of the queries.



Combining Searches

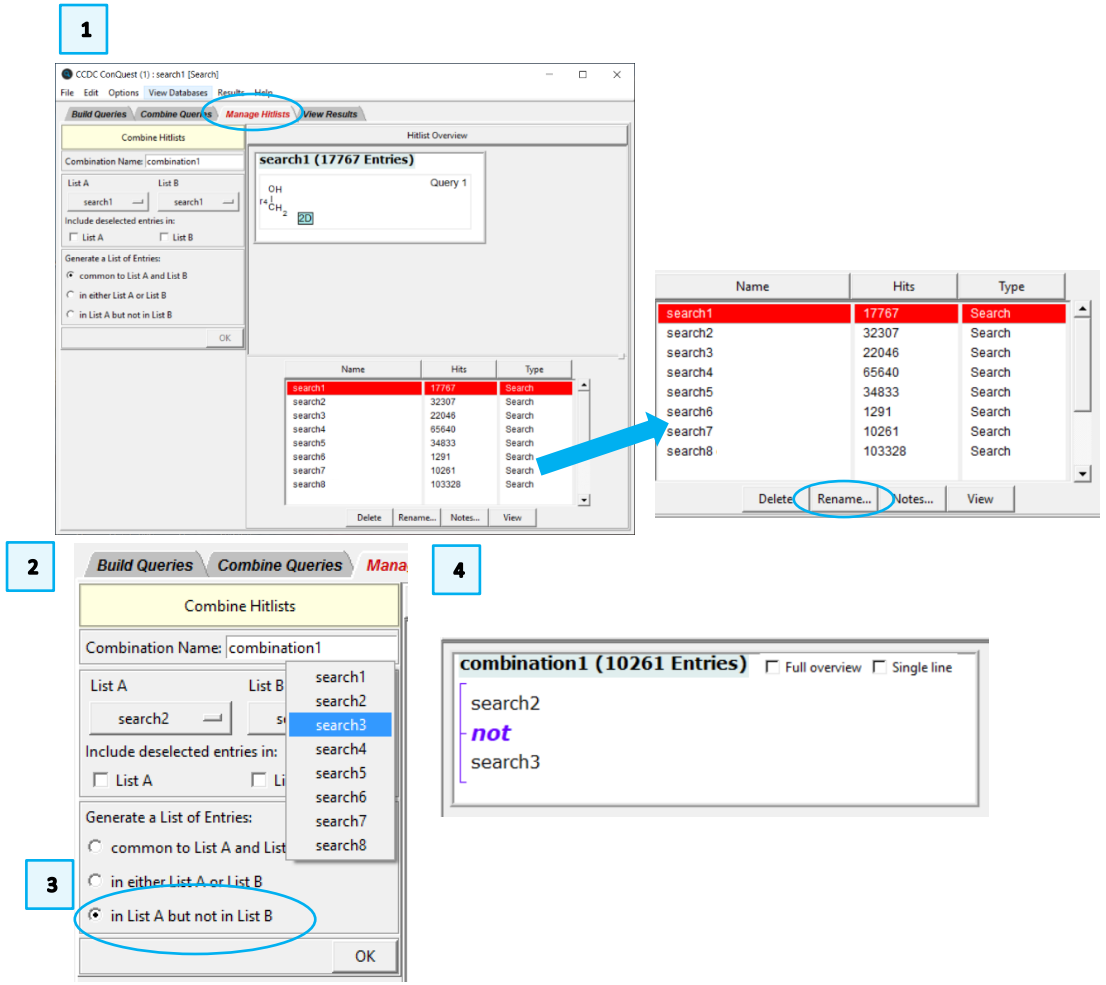
1. You can also combine search results using the *Manage Hitlist* tab. You will see a list of all the searches you performed. If it helps to keep them straight, you can select a search, click the **Rename…** button and enter a new name for the search.

2. If we wanted to generate the list of refcodes for the set of acyclic secondary alcohols, we could simply subtract the hits from search 3 (cyclic secondary alcohols) from the list of hits in search 2. To do this, in the drop-down box under List A, choose *search2*. Then in the box under List B choose *search3.*

3. Now we need to select the combination. For this we want the hits in search 2 (all secondary alcohols) but not the hits from search 3 (only cyclic alcohols). Tick the box for "in List A but not in List B".

4. Click **OK** and you will see a new item "combination 1" which shows how many hits are returned. You can see immediately here that this combination returns the same number of hits (at least 10261) as our previous Boolean combination of queries (search 7). Click the **View** button at the bottom of the window to view the highlighted combination or any of the searches.

5. You can also use these "combination" sets within new combinations to create very refined hitlists.



Exporting Results

1. In the *View Results* tab, you can choose which entries to save by toggling between the green check (selected) and the red cross (deselected).

2. To export one or all of your hits in various file formats, from the *View Results* tab, choose “File” from the top menu, then *Export Entries as…*

3. This will bring up a window where you can choose your desired file format. Choose the appropriate format for your work.

4. Use the dialogue to choose to export the *Current entry only* or *All selected entries*.

5. Different file types ask for different options, select these if appropriate.

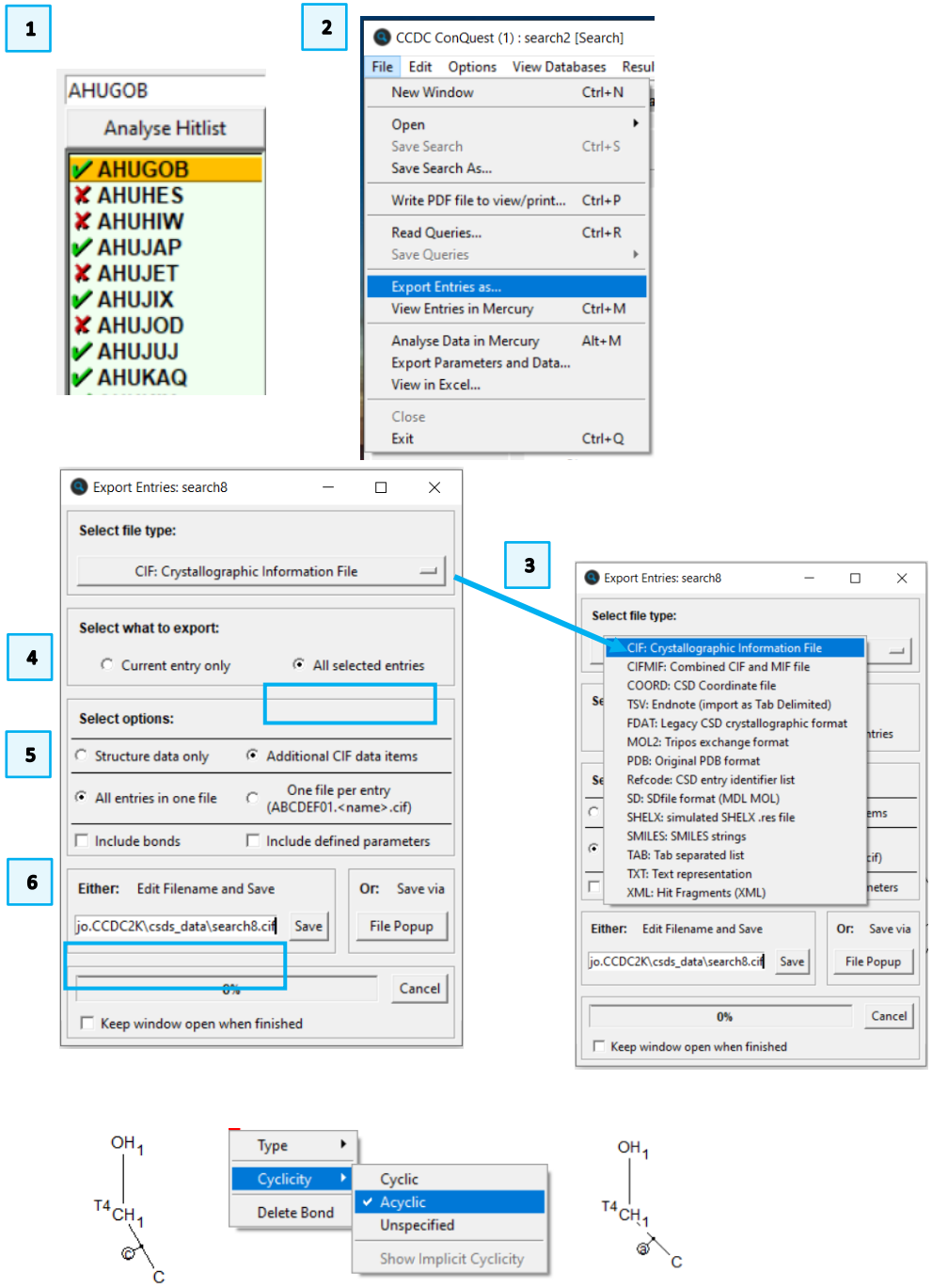
6. Either edit the file name and save, or choose Save via **File Popup** to choose the location manually.

Conclusions

In this example, you have seen how to make subtle changes to a search fragment and what effect those changes can have on the type and number of structures returned. You have also learned how to combine queries with Boolean operators, and how to create combinations of different searches using the hitlist manager. Finally, you learned how to export your results as different file types for use in other applications.

Exercise

- Try creating a search fragment that would return all the acyclic secondary alcohols without using combinations. ***Hint: Start with fragment c.***



Summary

This workshop introduced searching in ConQuest. You should now be familiar with:

- Sketching substructures using the ConQuest *Draw* window.
- Changing bond and atom types.
- How to use ring templates.
- Searching fragments and what effect slight changes can have on the hitlists returned.
- How to combine queries with Boolean operators.
- How to view the data for each hit in your results list.
- How to create combinations of different searches using the hitlist manager.
- How to export your results as different file types for use in other applications.

Next Steps

Advanced ConQuest workshops can be found here (https://www.ccdc.cam.ac.uk/community/training-and-learning/workshop-materials/csd-core-workshops/). The ConQuest user guide and other available ConQuest documents can be found here.

Feedback

We hope this workshop improved your understanding of ConQuest and you found it useful for your work. As we aim to continuously improve our training materials, we would love to get your feedback. Click on this link (https://www.surveymonkey.co.uk/r/CCDC-Online-Workshop) to a survey, it will take less than **5** minutes to complete. The feedback is anonymous. You will be asked to insert the workshop code, which for this self-guided workshop is CQ-001. Thank you!

Glossary

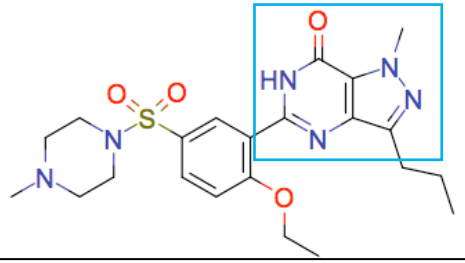
CSD Refcode - A database reference code, containing 6-8 characters and used to identify entries in the CSD. Learn more about CSD Refcodes in the blog: https://www.ccdc.cam.ac.uk/Community/blog/post-20/ .

Refcode list or ***.gcd* file** – A file containing a list of CSD Refcodes. This file can be opened in various CCDC applications.

Disordered structures – Disordered structures display a lack of regularity. For example, each of the F atoms in atrifluoromethyl group, -CF_3, might be randomly distributed between two sites. This means that the crystallographer will report two sets of coordinates for each F atom. In some cases, two alternative sites are occupied equally; in other cases, there is a major site and a minor site. Disorder can involve more than two sites and it can also involve a whole molecule. Disordered structures in the Cambridge Structural Database may have been treated in one of two ways. In older structures, only one position is kept for each disordered atom. When such a structure is read into Mercury, it will therefore appear as if it is not disordered at all. In more recent structures, all positions of disordered atoms are kept but only one set is connected by bonds to form a complete molecule. When a structure like this is displayed in Mercury, it will have the appearance of a complete molecule with disconnected atoms at various points to indicate alternative sites for the disordered atoms.

Substructure – A substructure is a part or section of a whole molecule.

Hitlist - A hitlist is a subset of CSD entries which can include search results, refcode lists, or the results of combining these.



Substructure used in example 1 search highlighted in the blue box.